7Б

УДК 004.08

В.П. Бурдаев
Харьковский национальный экономический университет, г. Харьков, Украина
burdaevvp@mail.ru

Об одном подходе реализации онтологии предметной области

Рассматривается модель онтологии, реализованная в почти интеллектуальной обучающей системе «КАРКАС». Анализируются ее возможности на примере определения экспозиции при оценке экологического риска.

Введение

Последние десятилетия особый интерес у исследователей искусственного интеллекта (ИИ) вызывают онтологии. Онтологии могут использоваться для представления знаний о понятиях предметной области (ПрО) и отношениях между ними, а также для описания содержания Web-страниц. Кроме того, онтологии можно использовать при построении баз знаний (БЗ) не только экспертных систем (ЭС), но также любых других приложений.

Одна из причин возникновения потребности в онтологии – это совместное ее использование как людьми, так и программными агентами. Например, консорциум W3C разрабатывает RDF (Resource Description Framework), язык кодирования знаний на Web-страницах, для того, чтобы сделать их понятными для программных агентов, которые осуществляют поиск информации. В этом смысле онтологии представляют собой интеллектуальные средства для развития и совершенствования сети Интернет.

Четкой границей между БЗ и онтологией предметной области не существует.

Целью данной работы является анализ реализации модели онтологии для почти интеллектуальной обучающей системы «КАРКАС» [1].

Модель онтологии

Построение модели ПрО основывается на понятии онтологии (наука о природе вещей и взаимосвязи между ними). С позиций ИИ термином «онтология» можно определить некоторый механизм, способ, который используется для описания ПрО, в частности базовых понятий этой области, их свойств и связей между ними. Другими словами, онтология состоит из объектов ПрО, разбитых на кластеры в соответствии с некоторыми критериями, а взаимодействие объектов выражается набором правил вывода. Современные онтологии кодируются таким образом, чтобы объекты, их свойства, правила вывода могли обрабатывать программные агенты [2].

Модель ПрО – это часть реального мира, отображаемая как класс или совокупность классов реальных объектов.

В ПрО выделяют категории:

– сущности – это объекты;
– отношения – это связи между объектами;
– атрибуты – это характеристики объектов.

Класс – это совокупность похожих объектов, которые имеют один или несколько атрибутов. Экземпляр класса – это объект, однозначно определяющийся значениями атрибутов.

Каждый класс может обладать любым количеством отношений с другими классами. Класс называется независимым, если каждый экземпляр его может быть однозначно идентифицирован без определения его отношений с другими классами. Класс называется зависимым, если его содержимое экземпляра зависит от его отношения к другому классу. Мощность класса – это количество его экземпляров.

Отношение связи (родитель-потомок) – это связь между классами, при которой каждый экземпляр класса-родителя ассоциирован с произвольным количеством экземпляров класса-потомка, а каждый экземпляр класса-потомка ассоциирован в точности с одним экземпляром класса-родителя.

Атрибут – это характеристика, описывающая что-либо в объекте. Каждому атрибуту присваивается уникальное имя, обозначающее его смысл и значение. Атрибут может иметь список возможных значений. Объект может обладать любым количеством атрибутов.

С позиций объектно-ориентированного программирования базовым понятием в онтологии является класс, который характеризуется свойствами и методами. Свойства класса задаются значениями его полей, а методы решают определенные задачи. Следовательно, онтология – это совокупность взаимодействующих объектов.

Наличие экземпляров классов (объектов), атрибутов и правил вывода в онтологии превращают ее из концептуальной схемы ПрО в БЗ.

Знания в онтологии могут быть представлены в виде логики предикатов, продукций, фреймов, семантической сети или каким-то другим способом (OWL – язык Web-технологий).

Одна из проблем онтологии Semantic Web состоит в том, что на настоящий момент времени существуют языки для описания онтологии, но нет БЗ, то есть экземпляров классов, объектов и логического вывода, другими словами, нет Web-знаний. Хотя заметим, что в Web 2.0 пользователи эффективно используют Web-сервисы, которые могут наполнить БЗ для Semantic Web.

Практически все модели онтологий содержат концепты (сущности, понятия, классы, объекты), свойства концептов (атрибуты, слоты), отношения между концептами (связи, зависимости).

Итак, онтология состоит из словаря и набора утверждений на некотором языке логики и является основой для коммуникации между людьми и программными агентами.

Онтология вместе с множеством экземпляров составляют БЗ и в этом смысле между онтологией и БЗ нет четкого разделения.

Средства построения онтологии в системе «КАРКАС». Предметом модели онтологии является абстрагирование сущностей реального мира в рамках ПрО, в результате которого выявляются объекты ПрО. Как правило, они обозначаются именем существительным естественного языка [3].

Объект описывается с помощью данных, именуемых свойствами или атрибутами объекта. Атрибуты являются определениями в высказывании об объекте и обозначаются именами существительными естественного языка. Взаимодействие объектов осуществляется через правила вывода, которые содержат атрибуты и их значения. Экземпляр объекта – это совокупность атрибутов и их значений, описывающих однозначно состояние объекта.

Задача выбора идентификатора объекта является семантически субъективной задачей. Поскольку объект определяется набором своих атрибутов, то для каждого объекта целесообразно выделить такое подмножество атрибутов, которое однозначно идентифицирует данный объект.

Идентификация объектов проводится когнитологами (аналитиками ПрО).

Различают однозначные и многозначные атрибуты. Однозначными являются атрибуты, которые в пределах конкретного экземпляра объекта имеют только одно значение. В противном случае они считаются многозначными.

Каждый атрибут имеет домен (domain). Домен – это выражение, определяющее значения, разрешенные для данного атрибута. Другими словами, домен – это область значений атрибута.

Когнитолог БЗ должен проконтролировать, чтобы в модели БЗ для каждого атрибута объекта был определен домен.

Итак, объекты не существуют отдельно друг от друга. Между ними имеются связи, которые отражаются в онтологии ПрО. Каждая связь реализуется через значения атрибутов объектов. Однотипные объекты объединяются в классы.

Иерархия классов в онтологии строится путем выделения так называемого базового класса, располагающегося на самом верхнем уровне. Далее выделяются подклассы, находящиеся на следующем уровне и так далее. Базовый класс можно рассматривать как логическую конструкцию модели для представления множества объектов и связей между ними как единого целого. Некоторую аналогию можно привести с построением расслоения. Тройка объектов (M, p, B) образует расслоение, где $p: M \rightarrow B$ – непрерывное отображение пространства расслоения M на пространство B (база расслоения), $X_b=p^{-1}(b)$ – слой расслоения. Типичным примером расслоения является касательное расслоение.

База расслоения содержит как базовый класс, так и его классы-потомки. Слой над каждым классом состоит из его экземпляров (рис. 1). Фазовый портрет динамической системы базы расслоения описывает связи между классами. Например, на рис. 1 базовый класс представляет собой репеллер (to repel – отталкивать) [4].

Простейший пример расслоения онтологии в обучении можно привести как конструкцию прямого произведения множества знаний преподавателей и множества знаний студентов с проекцией на множество знаний преподавателей. Тогда знания каждого преподавателя можно рассматривать как класс со своими свойствами методики преподавания дисциплины и методами обучения. В этом случае над точкой базы расслоения (знания преподавателя) будет находиться слой, состоящий из знаний студентов, которые обучаются дисциплине в соответствии со структурой класса базы расслоения. Другими словами, к знаниям студентов слоя применимы методы обучения, составленные преподавателем, в виде правил и фреймов.

Если рассматривать движение динамической системы в базе расслоения, то базовый класс и его подклассы будут перемещаться, что влечет за собой использование дополнительного параметра времени в правилах и фреймах. Следовательно, для отслеживания движения в расслоенном пространстве используются программные агенты. А характер движения знаний студентов (динамическая система в пространстве расслоения M) приобретает зависимость от движения знаний преподавателя в базе расслоения. Например, если преподаватель креативно (способность к обучению) совершенствует свои знания во время изложения дисциплины в течение нескольких лет (движение динамической системы базы расслоения приобретает характер возвращаемости по времени), то знания для студентов в расслоенном пространстве также становятся креативными. Естественно возникает актуальная задача исследования условий, обеспечивающих креативность знаний студентов.

Процедура построения онтологии состоит из следующих этапов:

– определение терминов, объектов;
– выделение классов (кластеризация) и их иерархия (базовый класс и их подклассы);
– создание связей между выделенными классами и объектами.

Классы и подклассы объектов определяются непосредственно в правилах и фреймах логического вывода при конструировании базы знаний.

Физическая модель БЗ хранит экземпляры классов, объектов, значения атрибутов объектов и логические связи между классами, объектами.

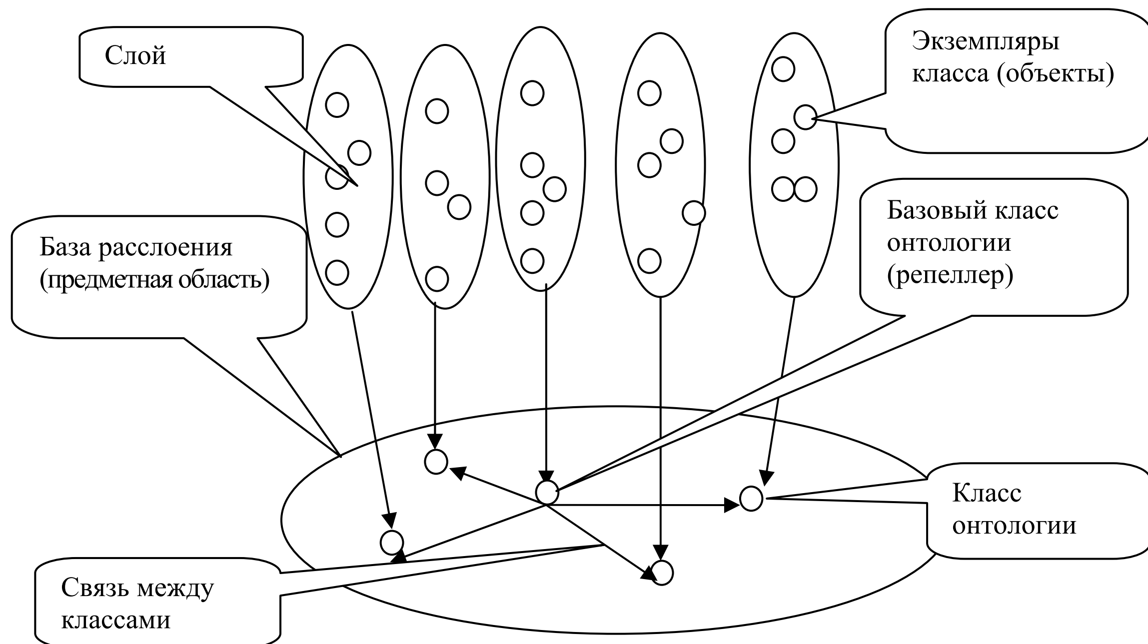


Рисунок 1 – Расслоение онтологии

Онтология в системе «КАРКАС» – это конструкция для агрегирования, множественных иерархий классов и правил вывода. В модели онтологии используются два класса, реализованные в виде продукции и фрейма. Класс рассматривается как шаблон, на основе которого создаются его экземпляры.

Разработка онтологий для «КАРКАС» состоит из 5 шагов:

– определение границ онтологии;
– определение доменов атрибутов ПрО;
– определение классов, объектов с помощью продукций и фреймов;
– кластеризация объектов и организация их иерархии;
– формирование продукций и фреймов для описания классов, подклассов, экземпляров.

Напомним, что онтология – это взгляд аналитика на ПрО, то есть всегда субъективна. Разработка онтологии включает: спецификацию, концептуализацию, формализацию и реализацию. Стадии формализации и реализации превращают онтологию в БЗ.

Спецификация. Вначале строится глоссарий терминов, включающий все термины (концепты и их экземпляры, атрибуты, действия и тому подобное), важные для ПрО, и их естественно-языковые описания.

Концептуализация. Определяются объекты (целевые, промежуточные и так далее) и их иерархия. Модуль кластеризации «КАРКАС» автоматически выполняет кластеризацию объектов. Таким образом, идентифицируются основные кластеры ПрО. Следующим шагом является построение дерева объектов онтологии, которое фиксирует связи между объектами онтологии.

Формализация. Определяется, какие классы системы «КАРКАС» будут использованы для онтологии предметной области.

Реализация. Формируется словарь атрибутов с соответствующими доменами (атрибуты и их значения) и экземпляры классов, используя шаблоны продукции и фрейма. Осуществляется тестирование БЗ (выявление конфликтов, противоречий в базе знаний и анализ ее полноты).

Пример модели онтологии ПрО для определения экспозиции при оценке экологического риска.

Переход от эмпирических оценок к методам принятия экологически верных решений лежит через создание ЭС экологического мониторинга, ориентированных на оценку и прогноз состояния окружающей природной среды, находящейся под антропогенным воздействием.

Окружающая природная среда (ОПС) рассматривается как неавтономная динамическая система с позиций понятия расширения динамических систем [4]. Фазовое пространство неавтономной динамической системы характеризуется пространством состояний, рассматриваемых как точка в некоторых обобщенных координатах, и задается системой дифференциальных уравнений (1).

$$\begin{cases} dx/dt = f(x,y) \\ dy/dt = g(y) \end{cases}. \quad (1)$$

Каждому состоянию системы, другими словами ее движению, отвечает определенная траектория перемещения отображающей точки в фазовом пространстве. Вопрос структурной устойчивости (грубости) таких систем рассмотрен в [4].

Для моделей онтологий с их бедной математической структурой (нет ни топологической, ни дифференцируемой структур) вопросы грубости систем не рассматриваются. Поэтому для конструирования онтологии ОПС необходимо использовать данные, которые относятся к определенному промежутку времени и определенной территории.

Постановка задачи. Создать онтологию по определению вида экспозиции при оценке экологического риска, используя анализ данных по каждой из видов экспозиции.

Назначение прототипа ЭС: это консультирование по определению вида экспозиции при оценке экологического риска (риски здоровью, риск смерти, риск заболевания).

Сфера применения прототипа ЭС: это различные муниципальные органы по оценке экологического риска.

Класс задач: это анализ возможных видов экспозиции.

Цель прототипа ЭС: это определение вида экспозиции при оценке экологического риска.

Ожидаемые результаты (список возможных значений цели консультации): наиболее правильное определение вида экспозиции при оценке экологического риска. Предполагаются следующие возможные значения цели консультации: ингаляционное поступление химических веществ (ХВ) от испарения из воды во время купания в открытом водоеме; ингаляционное воздействие ХВ, попадающих в воздух из почвы; ингаляционное поступление ХВ от испарения питьевой воды; пероральное поступление ХВ с питьевой водой; пероральное поступление ХВ при случайном заглатывании поверхностной воды; пероральное поступление ХВ из почвы; накожная экспозиция почвы; накожная экспозиция открытых водоемов; накожная экспозиция водопроводной воды.

Подцели: подцелями консультации являются среда распространения ХВ, пограничные органы человека, частота воздействия ХВ, продолжительность воздействия ХВ, агрегатное состояние ХВ.

Исходные данные:

– для анализа среды распространения – это идентификация той среды, которая переносит загрязняющее вещество (ЗВ);

– для определения способа поступления вещества в организм – это информация об агрегатном состоянии веществ, пограничных органах человека;

– для определения степени экспозиции – это информация о частоте и продолжительности воздействия.

Концептуальная модель ПрО. Экспозиция (воздействие) – это контакт организма (рецептора) с химическим, физическим или биологическим агентом. Сценарий экспозиции включает оценку поступления ХВ в организм человека одновременно из разных

сред: атмосферный воздух, питьевая вода, вода поверхностного водоема, почва, продукты питания – и различными путями: пероральный, ингаляционный, накожный.

Для определения вида экспозиции были выбраны следующие критерии:

– агрегатное состояние загрязняющего химического вещества;
– пограничные органы человека;
– частота воздействия;
– продолжительность воздействия;
– среда распространения загрязняющего вещества.

Соответственно, классы ПрО представлены в табл. 1.

Таблица 1 – Классы БЗ

Класс	Отношение связи	Мощность	Уровень иерархии
Вид экспозиции	Родитель	10	1
Агрегатное состояние	Потомок	6	2
Пограничные органы человека	Потомок	6	2
Частота воздействия	Потомок	4	2
Продолжительность воздействия	Потомок	6	2
Среда распространения	Потомок	6	2

В соответствии с методикой создания онтологии в системе «КАРКАС» первым шагом является ввод имен атрибутов (терминов), их описаний и доменов предметной области с помощью редактора БЗ.

Следующим шагом является формирование классов и их экземпляров (объектов) с помощью правил и фреймов. Редактор БЗ позволяет просмотреть и изменить порядок следования правил, фреймов и выполнить кластеризацию правил по классу верхнего уровня. На рис. 2 представлены как исходное расположение правил, так и результат кластеризации правил БЗ.

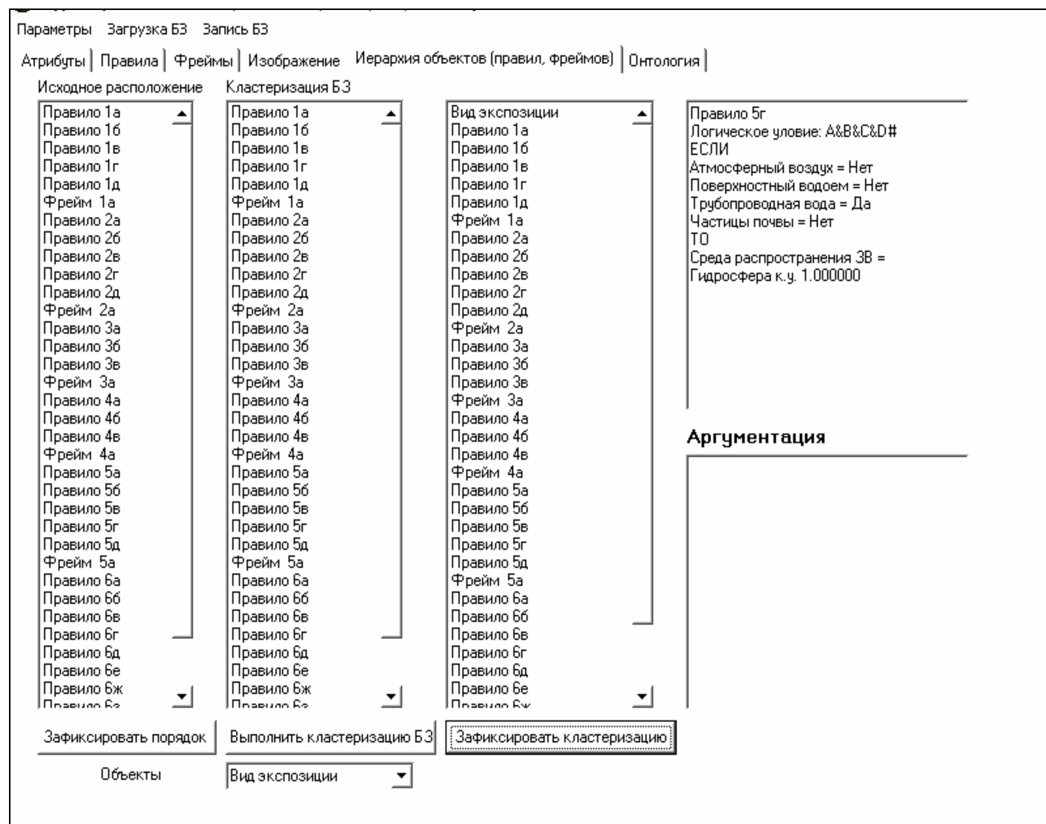


Рисунок 2 – Редактор БЗ. Порядок правил и фреймов в БЗ

Фреймы для данной модели играют роль аттракторов, которые улавливают не существующие экземпляры классов, образованные в результате консультации пользователя с системой.

Третий шаг предназначен для тестирования БЗ. Когнитолог выбирает цель консультации, объект: «Вид экспозиции» и режим консультации с объяснениями. Далее с помощью опции пункта меню «Начать» запускается машина логического вывода (по умолчанию активизируется машина вывода, реализующая «обратный вывод»), и когнитологу предлагается выбрать или ввести ответы на вопросы, которые генерирует машина вывода. В результате работы машины логического вывода формируются факты, которые можно просмотреть с помощью модуля базы фактов.

Последняя вкладка в редакторе БЗ дает возможность построить иерархию классов БЗ (рис. 3).

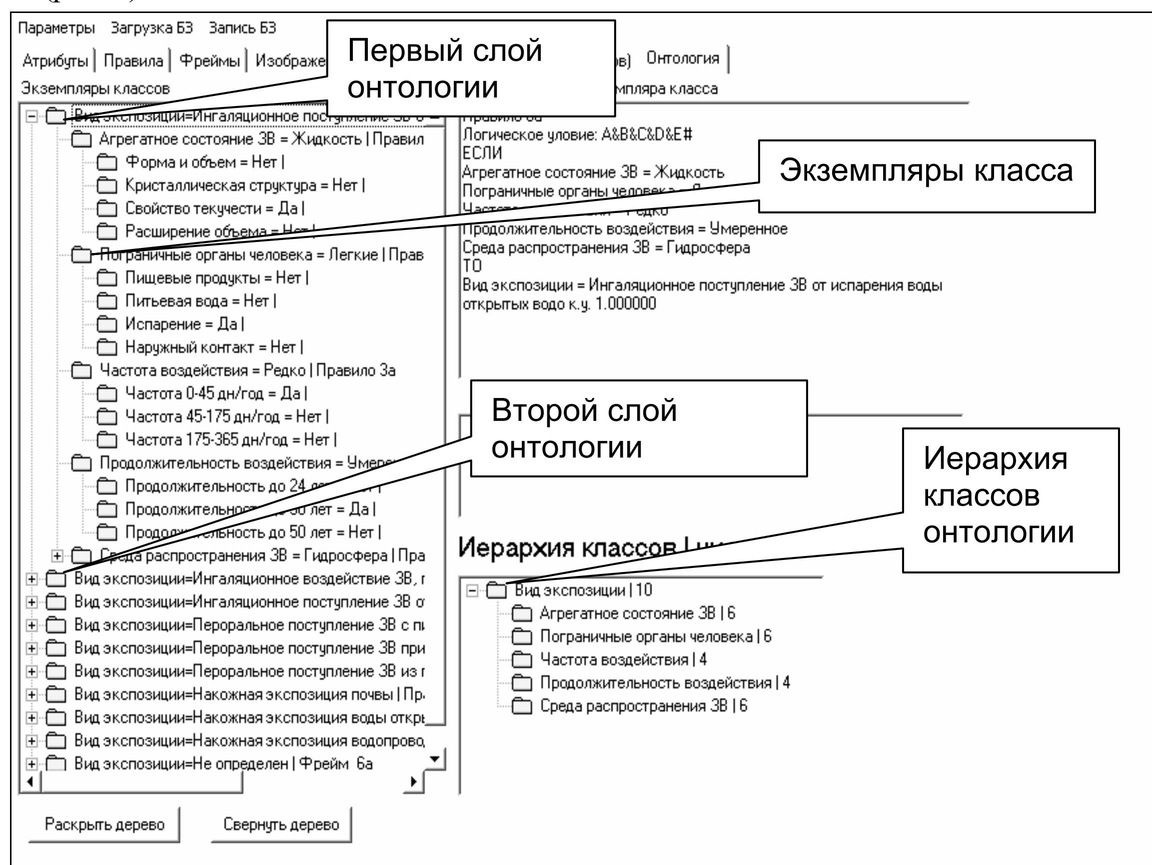


Рисунок 3 – Редактор БЗ. Онтология ПрО

Ниже приведен фрагмент БЗ для определения подцелей и целей консультации.

Экземпляры класса «Агрегатное состояние ЗВ».

Правило 1а. A&B&C&D#.
ЕСЛИ
A Форма и объем = Да
B Кристаллическая структура = Да
C Свойство текучести = Нет
D Расширение объема = Нет
ТО
Агрегатное состояние ЗВ = Твердое вещество.
Правило 1б. A&B&C&D#.
ЕСЛИ
A Форма и объем = Да

B Кристаллическая структура = Нет
C Свойство текучести = Нет
D Расширение объема = Нет
ТО
Агрегатное состояние ЗВ = Твердое вещество.
Правило 1в. A&B&C&D#.
ЕСЛИ
A Форма и объем = Нет
B Кристаллическая структура = Да
C Свойство текучести = Нет
D Расширение объема = Нет
ТО
Агрегатное состояние ЗВ = Твердое вещество.
Фрейм 1а.
Имя слота | Тип слота | Наследование
Форма и объем | Замещение | н
Кристаллическая структура | Замещение | н
Свойство текучести | Замещение | н
Расширение объема | Замещение | н
Целевой слот.
Агрегатное состояние ЗВ | Не определено.
Экземпляры класса «Пограничные органы человека».
Правило 2г. A&B&C&D#.
ЕСЛИ
A Пищевые продукты = Нет
B Питьевая вода = Нет
C Испарение = Да
D Наружный контакт = Нет
ТО
Пограничные органы человека = Легкие.
Правило 2д. A&B&C&D#.
ЕСЛИ
A Пищевые продукты = Нет
B Питьевая вода = Нет
C Испарение = Нет
D Наружный контакт = Да
ТО
Пограничные органы человека = Кожа.
Фрейм 2а.
Имя слота | Тип слота | Наследование
Пищевые продукты | Замещение | н
Питьевая вода | Замещение | н
Испарение | Замещение | н
Наружный контакт | Замещение | н
Целевой слот.
Пограничные органы человека | Не определены.
Экземпляры класса «Частота воздействия».
Правило 3а. A&B&C#.
ЕСЛИ
A Частота 0-45 дн./год = Да
B Частота 45-175 дн./год = Нет
C Частота 175-365 дн./год = Нет
ТО
Частота воздействия = Редко.

Фрейм 3а.
Имя слота | Тип слота | Наследование
Частота 0-45 дн./год | Замещение | н
Частота 45-175 дн./год | Замещение | н
Частота 175-365 дн./год | Замещение | н
Целевой слот.
Частота воздействия | Не определена.
Элементы класса «Продолжительность воздействия».
Правило 4а. A&B&C#.
ЕСЛИ
A Продолжительность до 24 лет = Да
B Продолжительность до 30 лет = Нет
C Продолжительность до 50 лет = Нет
ТО
Продолжительность воздействия = Не продолжительное.
Правило 4б. A&B&C#.
ЕСЛИ
A Продолжительность до 24 лет = Нет
B Продолжительность до 30 лет = Да
C Продолжительность до 50 лет = Нет
ТО
Продолжительность воздействия = Умеренное.
Экземпляры класса «Среда распространения ЗВ».
Правило 5а. A&B&C&D#.
ЕСЛИ
A Атмосферный воздух = Да
B Поверхностный водоем = Нет
C Трубопроводная вода = Нет
D Частицы почвы = Нет
ТО
Среда распространения ЗВ = Атмосфера.
Правило 5б. A&B&C&D#.
ЕСЛИ
A Атмосферный воздух = Нет
B Поверхностный водоем = Да
C Трубопроводная вода = Да
D Частицы почвы = Нет
ТО
Среда распространения ЗВ = Гидросфера.
Фрейм 5а.
Имя слота | Тип слота | Наследование
Атмосферный воздух | Замещение | н
Трубопроводная вода | Замещение | н
Поверхностный водоем | Замещение | н
Частицы почвы | Замещение | н
Целевой слот.
Среда распространения ЗВ | Не определена.
Экземпляры класса «Вид экспозиции».
Правило 6а. A&B&C&D&E#.
ЕСЛИ
A Агрегатное состояние ЗВ = Жидкость
B Пограничные органы человека = Легкие
C Частота воздействия = Редко
D Продолжительность воздействия = Умеренное

Е Среда распространения ЗВ = Гидросфера
ТО
Вид экспозиции = Ингаляционное поступление ЗВ от испарения воды открытых водоемов.
Правило 6б. A&B&C&D&E#.
ЕСЛИ
А Агрегатное состояние ЗВ = Газообразное вещество
В Пограничные органы человека = Легкие
С Частота воздействия = Очень часто
D Продолжительность воздействия = Длительное
Е Среда распространения ЗВ = Атмосфера
ТО
Вид экспозиции = Ингаляционное воздействие ЗВ, попадающих в воздух из почвы.

Итак, онтология ПрО содержит 5 классов и двухуровневую структуру классов.

Выводы

Неформально онтология состоит из терминов и правил использования этих терминов. На формальном же уровне онтология – это система, состоящая из набора понятий и связей между ними, на основе которых можно строить классы, объекты, отношения, функции и теории. Частными случаями онтологий являются простой словарь, тезаурус.

Модель онтологии предметной области в системе «КАРКАС» состоит из иерархии понятий предметной области, связей между ними (правил вывода), которые действуют в рамках этой модели. Предложен механизм интерпретации онтологии в условиях динамического изменения ее параметров (базового класса, связей между классами и взаимодействия объектов классов). Заметим, что в системе Protégé [5] все понятия предметной области делятся на классы, подклассы, экземпляры. Экземпляры могут быть как у класса, так и подкласса, и описываются они фреймом.

Однако самые изощренные редакторы баз знаний и инструменты онтологий не могут выполнить содержательный, креативный анализ предметной области для создания онтологии без вмешательства когнитолога.

Литература

1. Бурдаев В.П. ПИОС – почти интеллектуальная обучающая система / В.П. Бурдаев // Искусственный интеллект. – 2009. – № 3. – С. 205-212.
2. Бурдаев В.П. Мультиагентная система в обучении / В.П. Бурдаев // Тезисы VII Международной конференции ИМС'2006. – Таганрог : Изд. ТРТУ, 2006. – С. 182-185.
3. Бурдаєв В.П. Системи навчання з елементами штучного інтелекту / Бурдаєв В.П. – Харків : Вид. ХНЕУ, 2010. – 400 с.
4. Бронштейн И.У. Неавтономные динамические системы / Бронштейн И.У. – Кишинев : Штиинца, 1984. – 290 с.
5. Режим доступа: http://protege.stanford.edu/

В.П. Бурдаєв

Про один підхід реалізації онтології предметної області
Розглядається модель онтології, реалізована в майже інтелектуальній навчальній системі «КАРКАС». Аналізуються її можливості на прикладі визначення експозиції при оцінці екологічного ризику.

V.P. Burdayev

About One Approach of Realization Ontologia Subject Domain
The model ontologia, realized in almost intellectual training system «KARKAS» is considered. It is analyzed its opportunities by the example of definition of an exposition at an estimation of ecological risk.

Статья поступила в редакцию 24.06.2010.